

CLAIMS

We claim:

1. A method for identifying genomic sequences unique to a set of organisms, the method comprising:
 - obtaining genomic data characteristic of the set;
 - formatting the genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;
 - searching a selected genomic database using the query and the similarity search engine; and
 - parsing the results of the search for those sequences showing uniqueness to the set.
2. The method of Claim 1, wherein the similarity search engine is a BLAST search engine.
3. The method of Claim 2, wherein the selected database is GenBank.
4. A computer program for identifying genomic sequences unique to a set of organisms, the computer program product comprising:
 - a computer-readable medium;
 - a genomic data interface module, stored on the medium and operable to couple to a source of genomic data to receive genomic data characteristic of the set;
 - a formatting module, stored on the medium and operable to format received genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;
 - a search interface module, stored on the medium and operable to interface with the similarity search engine to submit the query-length sequence to the search engine
 - a search results parsing module, stored on the medium and operable to parse the results of the search for those sequences showing uniqueness to the set.

5. A method for identifying oligonucleotide sequences unique to a set of organisms, the method comprising:

- obtaining genomic data characteristic of the set;
- first formatting the genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a first similarity search engine;
- first searching a first selected genomic database using the query-length sequence and the first similarity search engine;
- first parsing the results of the first search for those genomic sequences showing uniqueness to the set;
- dividing at least one genomic sequence showing uniqueness to the set into a plurality of target-length oligonucleotide sequences;
- second formatting a plurality of the target oligonucleotide sequences into a query format compatible with a second similarity search engine;
- second searching a second selected genomic database using the formatted target oligonucleotide sequences and the second similarity search engine;
- second parsing the results of the second search for those oligonucleotides showing uniqueness to the set.

6. The method of Claim 5, wherein:

the first and second similarity search engines are BLAST search engines.

7. The method of Claim 5, wherein:

The selected database is GenBank.

8. A computer program for identifying oligonucleotide sequences unique to a set of organisms, the computer program product comprising:

- a computer-readable medium;
 - a genomic data interface module, stored on the medium and operable to couple to a source of genomic data to receive genomic data characteristic of the set;
 - a first formatting module, stored on the medium and operable to format received genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;
 - a first search interface module, stored on the medium and operable to interface with the similarity search engine to submit the query-length sequence to the search engine
 - a first search results parsing module, stored on the medium and operable to parse the results of the search for those sequences showing uniqueness to the set.
- dividing at least one genomic sequence showing uniqueness to the set into a plurality of target-length oligonucleotide sequences;
- a second formatting module, stored on the medium and operable to format oligonucleotide sequences into a format compatible with a similarity search engine;
 - a second search interface module, stored on the medium and operable to interface with the similarity search engine to submit the formatted oligonucleotide sequence to the search engine; and
 - a second search results parsing module, stored on the medium and operable to parse the results of the search for those oligonucleotide sequences showing uniqueness to the set.

9. The computer program product of Claim 8, wherein the first and second search modules are combined.

10. The computer program product of Claim 8, wherein the first and second parsing modules are combined.

11. The computer program product of Claim 8, wherein the first and second formatting modules are combined.

12. A method for inferring genomic sequences unique to a second set of organisms, the method comprising:

- obtaining genomic data characteristic of a first set of organisms;
- formatting the genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;
- searching a selected genomic database using the query and the similarity search engine;
- parsing the results of the search for those sequences not associated with the first set of organisms, but showing similarity beyond a threshold.

13. A computer program for inferring genomic sequences unique to a second set of organisms, the computer program product comprising:

- a computer-readable medium;
- a genomic data interface module, stored on the medium and operable to couple to a source of genomic data to receive genomic data characteristic of a first set of organisms;
- a formatting module, stored on the medium and operable to format received genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;
- a search interface module, stored on the medium and operable to interface with the similarity search engine to submit the query-length sequence to the search engine; and
- a search results parsing module, stored on the medium and operable to parse the results of the search for those sequences not associated with the first set of organisms, but showing similarity beyond a threshold.